# Integrated Global-Local Metric Learning for Person Re-identification

Jing Zhang Xu Zhao
Department of Automation, Shanghai Jiao Tong University, China
Key Laboratory of System Control and Information Processing, Ministry of Education, China
{crystal_zj, zhaoxu}@sjtu.edu.cn

## Abstract

*The task of person re-identification (re-id) is to match images of people observed in different camera views. Recent researches mainly focus on feature representation and metric learning. Many global metric learning approaches have achieved good performance. Since comparing all of the samples with a single global metric is inappropriate to handle heterogeneous data, some local metric learning approaches are proposed. But most of them cannot be used on re-id directly due to some research challenges. Also, they usually need complicated computation to solve the optimization problems with numerous parameters. In order to improve the performance of global metric learning and avoid complex computation, we propose to simultaneously learn local metrics on clusters of samples softly partitioned by Gaussian Mixture Model (GMM) and a global metric on the entire training set. Then the local metrics are combined with the global metric by their posterior probabilities of GMM to obtain an integrated metric for similarity evaluation. Experiments on three challenging datasets (VIPeR, PRID450S and QMUL GRID) verify the effectiveness of the proposed method.*

## 1. Introduction

Person re-identification (re-id) is an important issue in the area of intelligent surveillance. Its target is to match snapshots of people observed in non-overlapping camera views. A person might show different appearances in different views due to variations of illumination, poses, viewpoints, background environments and occlusions.

Recent works mainly focus on feature representation [2,3,4,14,15,34] and metric learning [2,5,6,16,17,22,26]. Some researchers try to propose feature descriptors which are discriminative to distinguish different persons and robust against intra-class variations of appearances. Local color and texture descriptors are generally used in feature representation. Many features achieve good performance, such as Ensemble of Local Features (ELF) [14], salience match [15], Weighted Histograms of Overlapping Stripes (WHOS) [34], Gaussian of Gaussian (GOG) [3], Local Maximal Occurrence (LOMO) [2] and an enhanced deep feature extracted by Feature Fusion Net (FFN) [4].

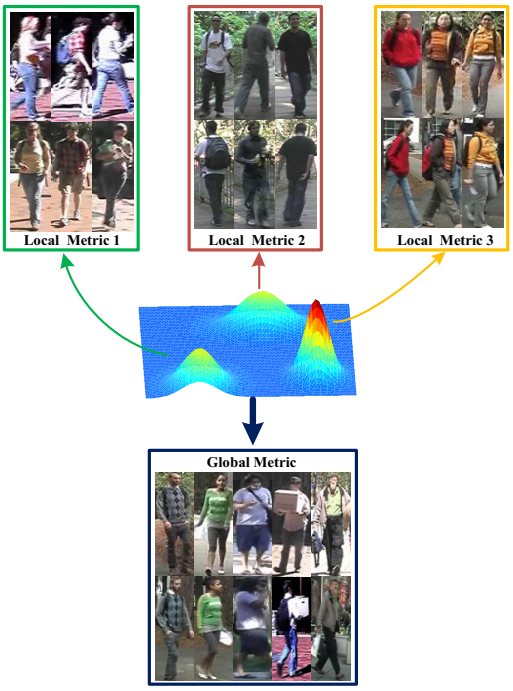


Figure 1: Comparison of global and local metric learning. Global metric is learned from the whole training set. Local metrics are learned from local clusters. There are sample images from VIPeR dataset [1] in each block, where images on the same column represent the same person under different views.

Metric learning is applied to learn a metric adapted to the features of training samples. Its goal is to ensure high similarity between intra-class samples and low similarity between inter-class samples. PRDC [16] maximizes the probability of that intra-class sample pairs have a relatively smaller distance than inter-class pairs. KISSME [5] and XQDA [2] fit the difference of sample pairs with Gaussian models. Then log ratio of the Gaussians is used to obtain a distance metric. MLAPG [6] minimize a logistic loss of training samples. These global metric learning approaches

978-1-5090-4822-9/17 $31.00 © 2017 IEEE
DOI 10.1109/WACV.2017.72

learn a single metric to measure all of the samples, which suffers limitations when handling datasets which vary locally.

Local metric learning [7,17,18,19,20,22] can emphasize local differences when comparing different samples. Bohné *et al.* [7] propose Large Margin Local Metric Learning (LMLML), which achieves good performance on face verification and handwritten digits classification problems. Li *et al.* [17] propose a local metric learning approach for re-id by jointly partitioning the training data according to the similarity of cross-view transforms.

Figure 1 shows the comparison of global and local metric learning in re-id. Images of re-id dataset usually vary locally in some clusters consisting of person images with similar clothes and background environments. Global metric is learned on the whole training set, which is more discriminative than the traditional Euclidean distance. However, it might neglect the locally subtle differences of similar images. Local metrics learned from local subsets focus more on local and individual differences of similar samples. Thus, combining global and local metrics might achieve better performance on re-id.

Motivated by the advantages of local and global metric learning, we propose a novel approach called Integrated Global and Local Metric Learning (IGLML). We utilize the idea of local metric learning and combine some recently proposed global metric learning methods such as XQDA [2] and MLAPG [6]. In the training stage, Gaussian Mixture Model (GMM) is used to cluster the training samples. Then a strategy for dividing the training set into several local subsets with overlaps is used. Local metrics are learned on each subset respectively. In the testing stage, for each pair of testing samples, the local metrics weighted by the GMM's posterior probabilities of each sample and the global metric weighted by a cross-validated parameter are summed up to obtain a final integrated metric. In this way, we use different metrics to compare different sample pairs. The proposed method improves the performance of metric learning by emphasizing more on local and individual differences. Additionally, local metrics are learned by softly integrating some global learning methods, which avoids complex computation of solving the optimization problems in many existing local metric learning methods.

The proposed approach is a general framework that integrates global and local metrics. It can be used to improve many existing global metric learning methods. We also realize the proposed framework with different feature representations and metric learning approaches. The results of experiments on three challenging datasets (VIPeR [1], PRID450S [10] and QUML GRID [11]) demonstrate the effectiveness of the proposed approach. And the results show that our work is generally effective with both different features and metric learning methods.

## 2. Related works

Metric learning plays an important role in re-id because it can obviously improve the performance even if the feature descriptors are not discriminative enough. Its goal is to learn a metric which ensures that intra-class samples have higher similarity than inter-class samples. Global metric learning approaches [2,5,6,16,26] learn a single metric to measure all of the samples, while local metric learning approaches [7,17,18,19,20,22] learn series of local metrics which are combined into an adaptive distance function when comparing different samples. Next, we review some global and local metric learning approaches.

### 2.1. Global metric learning

Many global metric learning approaches have been proposed to solve re-id problems. The Mahalanobis distance is a generally used linear metric. For two samples, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, the metric is defined as $\left(\boldsymbol{x}_i-\boldsymbol{x}_j\right)^{\mathrm{T}}\boldsymbol{M}\left(\boldsymbol{x}_i-\boldsymbol{x}_j\right)$, where $\boldsymbol{M}$ is called metric matrix. XQDA [2] and MLAPG [6] are global metric learning approaches recently proposed.

**XQDA:** Cross-view Quadratic Discriminant Analysis (XQDA) [2] is proposed based on Keep It Simple and Straightforward Metric (KISSME) [5] and Bayesian Face [9] approaches. Gaussian model is used to fit the distributions of differences between intra-class and inter-class samples respectively. The Mahalanobis metric is derived by the log-likelihood ratio of the two Gaussians. For two samples, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, the derived distance function is

$$d^2\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)=\left(\boldsymbol{x}_i-\boldsymbol{x}_j\right)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_I^{-1}-\boldsymbol{\Sigma}_E^{-1}\right)\left(\boldsymbol{x}_i-\boldsymbol{x}_j\right) \quad (1)$$

where $\boldsymbol{\Sigma}_I$ and $\boldsymbol{\Sigma}_E$ are covariance matrixes of differences between intra-class and inter-class sample pairs respectively. XQDA learns a discriminant low dimensional subspace and a metric simultaneously. It reduces the dimensions of features considering the influence of dimension reduction on metric learning. For the original features, $\boldsymbol{x}_i,\boldsymbol{x}_j\in\mathrm{R}^d$, XQDA learns a matrix $\boldsymbol{W}\in\mathrm{R}^{d\times d'}(d'<d)$ to map the features to a lower dimensional subspace. Considering the mapping matrix $\boldsymbol{W}$, the distance function is defined as

$$d_{\boldsymbol{W}}^2\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)=\left(\boldsymbol{x}_i-\boldsymbol{x}_j\right)^{\mathrm{T}}\boldsymbol{W}\left(\boldsymbol{\Sigma}_I^{'-1}-\boldsymbol{\Sigma}_E^{'-1}\right)\boldsymbol{W}^{\mathrm{T}}\left(\boldsymbol{x}_i-\boldsymbol{x}_j\right) \quad (2)$$

where $\boldsymbol{\Sigma}_I'=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Sigma}_I\boldsymbol{W}$ and $\boldsymbol{\Sigma}_E'=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\Sigma}_E\boldsymbol{W}$.

**MLAPG:** MLAPG [6] learns a metric by minimizing a log logistic loss function on the entire training set,

$$F\left(\boldsymbol{M}\right)=\sum_{i=1}^{n}\sum_{j=1}^{m}w_{ij}f_{\boldsymbol{M}}\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right) \quad (3)$$

where $f_{\boldsymbol{M}}\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)$ is the log logistic loss of the sample pair

$(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $n$ and $m$ are the total numbers of samples from two camera views, $w_{ij}$ is used to balance the loss of inter-class and intra-class sample pairs. If $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are in the same class, they are positive pair and $w_{ij} = 1/N_+$, otherwise, $w_{ij} = 1/N_-$, where $N_+$ and $N_-$ are the total number of positive and negative sample pairs in the training set.

The optimization problem is described as

$$\min F(\boldsymbol{M}) \quad \text{s.t. } \boldsymbol{M} \succ = 0 \tag{4}$$

where $\boldsymbol{M} \succ = 0$ means that $\boldsymbol{M}$ is a positive semi-definite (PSD) matrix. The accelerated proximal gradient (APG) [28] approach is used to solve the equation (4).

### 2.2. Local metric learning

Many local metric learning approaches are proposed to obtain more flexible metrics in order to handle datasets with complex distributions. They usually learn different metrics on clusters of training data. These approaches achieve good performance on object classification and face recognition problems mainly. However, many existing local metric learning methods are not suitable to be used on re-id directly due to the following challenges: (1) Only few samples in each class of a person. Especially, there are only two samples for a person in the single-shot case of re-id. (2) Persons' ids in testing stage are never been seen in training stage. (3) Larger intra-class variations than classification and face recognition problems. (4) High-dimensional features to ensure the discriminative power.

LMLML [7] computes a set of local metrics by optimizing a convex problem which favors a large margin solution. When measuring the similarity of two samples, the learned local metrics are combined according to the softly partitioning of training data. LMLML achieves good performance on various datasets including handwritten digits and face verifications *et al*. However, it needs a complex procedure of solving convex optimization problems to get local metrics. The number of parameters is related to the dimensions of features. Due to the above challenge (4) of re-id, when processing high-dimensional features, it's difficult to solve the optimization problems. Additionally, Bohné *et al.* [7] mention that LMLML has limits on the datasets with wide intra-class variations. On account of the challenge (3), samples of the same person in re-id dataset might be partitioned into different local sets. It makes difficulties for localized learning.

Cluster-based Adaptive Metric (CLAM) [20] performs well on object classification problems. An iterative hierarchical clustering method is used on each class of the training data. Then each cluster is regarded as a separate class and modeled by a Gaussian distribution. The classification result is determined by the sum of Bayesian posterior probabilities of the clusters in each class. However, it's unrealistic to cluster data of each class of a person for re-id due to the above challenge (1) and (2).

Coordinated Local Metric Learning (CLML) [18] use GMM to obtain soft-partitioning of the data. The feature $\boldsymbol{x}_i$ is multiplied by $p(k \mid \boldsymbol{x}_i)$, which is the posterior probability of $\boldsymbol{x}_i$ assigning to cluster $k$, and it forms a high-dimensional feature $\boldsymbol{x}_i' = \left( p(1 \mid \boldsymbol{x}_i)(\boldsymbol{x}_i^{\mathrm{T}}, 1), ..., p(K \mid \boldsymbol{x}_i)(\boldsymbol{x}_i^{\mathrm{T}}, 1) \right)^{\mathrm{T}}$, where $K$ is the number of clusters. Then the existing global metric learning methods can be used on the new feature to obtain a metric. New feature $\boldsymbol{x}_i'$ owns $K$ times of dimensions than the original feature $\boldsymbol{x}_i$. As mentioned in the challenge (4), many existing global metric learning methods have limits when facing the extremely high-dimensional features.

Li *et al.* [17] divides the images of different views according to the similarity of cross-view transforms. And classifiers are learned locally on each divided set. Samples with similar transforms are projected to a common feature space and combined softly for matching. Instance Specific Distance (ISD) [24] even learns a different classifier for each training sample by metric propagation strategy.

Local metric learning approaches can adapt to the local characteristics of data. However, most of the existing local learning approaches cannot be used on re-id directly due to the four challenges mentioned above. We propose a novel framework of integrating global metric learning methods with the idea of localized learning. The performance of many existing metric learning approaches can be improved with the proposed framework, since it not only takes advantages of metric learning, but also focuses more on local and individual differences by integrating local metrics.

## 3. Integrated global-local metric learning

Motivated by the advantages of existing global metric learning approaches and the idea of local learning, we propose to integrate global and local metrics to obtain better performance. The local metrics are learned on each cluster softly partitioned by GMM [26], and the global metric is learned on the whole training set. Then the global and local metrics are combined adapted to individual and local characteristics when measuring similarity of sample pairs. We call the proposed framework as IGLML.

### 3.1. Model

Figure 2 and 3 show the training and testing procedures of the proposed approach respectively. In the training stage, the training data is clustered by GMM with $K$ components. Then for each cluster, a metric is learned separately. Thus, a series of local metrics, $\boldsymbol{M}_k (k = 1, 2, ..., K)$, are obtained. Meanwhile, the global metric matrix $\boldsymbol{M}_0$ is learned from all of the training samples.

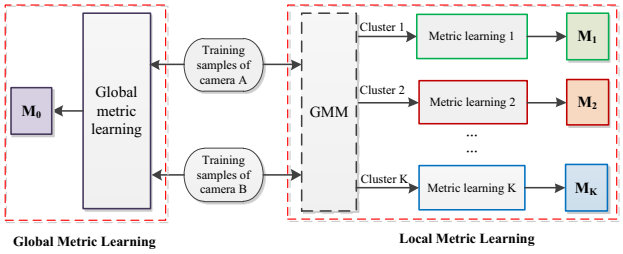


Figure 2: Training procedure of Integrated Global-Local Metric Learning (IGLML) method. Local metrics are learned on clusters of GMM. Global metric is learned on the entire training set.

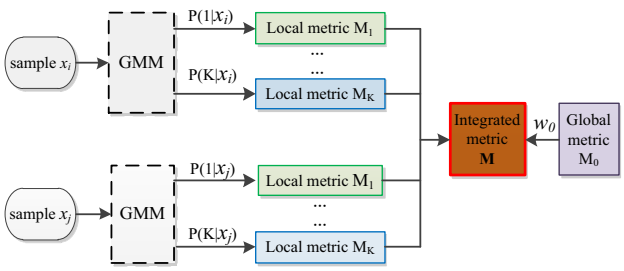


Figure 3: Testing procedure of Integrated Global-Local Metric Learning (IGLML) method. Local metrics are weighted by samples' posterior probabilities of Gaussian components. Local and global metrics are integrated as the final metric.

In the testing stage, we use the weighted sum of local and global metrics to compare sample pairs. The weight of each local metric is given by the samples' posterior probabilities aligning to the *k*-th GMM component. And the weight of the global metric $\boldsymbol{M}_0$ is given by cross-validation.

### 3.2. Local metric training

GMM is composed by multiple components of Gaussian distributions. It can be used for clustering and fitting complex distributions. A sample belonging to which cluster is determined by its posterior probabilities of each Gaussian component. GMM can be used to partition samples softly.

In the training stage, GMM with *K* components is used to fit the distribution of the entire training set. The training set is partitioned into several local subsets according to the posterior probability of each sample's alignment to each Gaussian component. Then metric learning approach is used on each local training set separately. Thus, a set of local metrics, $\boldsymbol{M}_k (k = 1,2,...,K)$, are obtained in this way.

Due to the greatly different appearances of the same person in different camera views, images with the same label might belong to different clusters of GMM. In order to ensure enough samples for training in each local training set

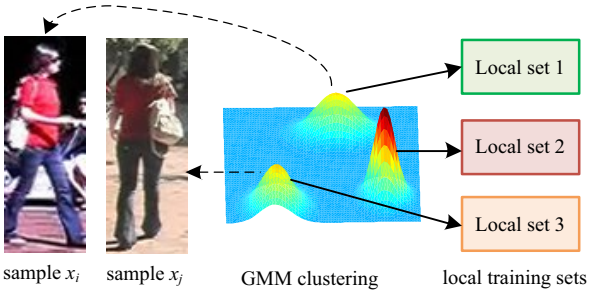


Figure 4: Sample $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ from the same class are partitioned to different clusters by GMM. In this situation, both $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are assigned to both the "Local set 1" and the "Local set 3" at the same time.

to avoid over-fitting, if a sample $\boldsymbol{x}_i$ belongs to the *k*-th cluster, samples having the same label with $\boldsymbol{x}_i$ are assigned to the same local training set. Thus, local metrics are training on overlapping local subsets. As Figure 4 shows, a pair of samples, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, own the same label. According to the maximum posterior probability of GMM components, $\boldsymbol{x}_i$ is assigned to the $k_i$ component corresponding to "Local set **1**", while $\boldsymbol{x}_j$ is assigned to another component $k_j$ ($k_i \neq k_j$) corresponding to "Local set **3**". In this situation, both $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are partitioned to both "Local set **1**" and the "Local set **3**" for local training simultaneously.

Features' dimensions for GMM unsupervised learning shouldn't be too large because the number of parameters in GMM is positively related to the features' dimensions. Estimating too many parameters would be prohibitive on account of the limited number of training samples for re-id. Moreover, the features' dimensions need to be reduced in order to ensure the smoothness of posterior probabilities of Gaussian components. Therefore, Principal Component Analysis (PCA) [21] is used to reduce the dimensions before clustering with GMM. The features don't need to be very discriminative because GMM is used in a very primary stage before metric learning. Reduction of dimensions has little negative effects on the final matching results.

### 3.3. Integrated global and local metric

In global metric learning approaches, Mahalanobis distance of a sample pair, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, is defined as

$$d^2(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{M}) = (\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} \boldsymbol{M} (\boldsymbol{x}_i - \boldsymbol{x}_j) \qquad (5)$$

In this work, we replace metric matrix $\boldsymbol{M}$ with a function of integrated global and local metrics, $\mathcal{M}(\boldsymbol{x}_i, \boldsymbol{x}_j)$, which is defined as

$$\mathcal{M}(\boldsymbol{x_i}, \boldsymbol{x_j}) = \alpha_0 \boldsymbol{M_0} + \sum_{k=1}^{K} \alpha_k \left(\boldsymbol{x_i}, \boldsymbol{x_j}\right) \boldsymbol{M_k} \quad (6)$$

where $\boldsymbol{M_0}$ is a global metric matrix learned on the entire training set. And $\boldsymbol{M_k}(k = 1, 2, ..., K)$ are a series of local metrics learned from each local training dataset. Their weights $\alpha_k(k = 1, 2, ..., K)$ are defined as

$$\alpha_k \left(\boldsymbol{x_i}, \boldsymbol{x_j}\right) = p\left(k \mid \boldsymbol{x_i}\right) + p\left(k \mid \boldsymbol{x_j}\right) \quad (7)$$

where $p\left(k \mid \boldsymbol{x_i}\right)$ is the posterior probability of that sample $\boldsymbol{x_i}$ is generated by the *k*-th component of GMM. Notice that $\sum_{k=1}^{K} \alpha_k \left(\boldsymbol{x_i}, \boldsymbol{x_j}\right) = 2$. $\boldsymbol{M_k}$ is a local metric learned on the *k*-th cluster of GMM, which have a greater effect in $\mathcal{M}(\boldsymbol{x_i}, \boldsymbol{x_j})$ if $\boldsymbol{x_i}$ or $\boldsymbol{x_j}$ is more strongly associated with *k*-th Gaussian component, and the effect is even greater if both $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ are more likely generated by the *k*-th component.

GMM tends to roughly partition images of persons with similar background environments or clothes into a local training set. Thus, when learning a local metric separately on each cluster, it will emphasize the local differences which are more discriminative between similar samples in the same local training set. And the local metrics $\boldsymbol{M_k}(k = 1, 2, ..., K)$ are combined into an adaptive metric for the final similarity computation. Theoretically, when the number of Gaussian components *K* is larger, the model is more able to learn a metric adapted to subtle local differences, but it's easier to be over-fitting. So the parameter *K* needs to be adjusted by cross-validation.

$\boldsymbol{M_0}$ is the global metric learned from the entire training dataset. $\alpha_0$ is a weight for balancing the influence of local and global metrics in the final integrated metric. If $\alpha_0 = 0$, the global metric has no effect on the integrated metric. The experimental results in section 4.3 show that purely local metric learning performs a little better than or equally to the corresponding global approach when $\alpha_0 = 0$. And the accuracy achieves a further improvement by adjusting the weight of the global metric $\alpha_0$.

### 3.4. Local metric learning for MLAPG

As described in the related work in section 2.1, we notice that MLAPG [6] use an asymmetric sample weighting strategy. In the equation (3), $w_{ij}$ is the weight of the loss of a sample pair $\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right)$. It motivates us to improve MLAPG with local metric learning idea further more.

Considering learning a local metric $\boldsymbol{M_k}$ on the local training set *k*, we replace the weight $w_{ij}$ with $w_{ij}^k$, which is defined as

$$w_{ij}^k = w_{ij}\left(p\left(k \mid \boldsymbol{x_i}\right) + p\left(k \mid \boldsymbol{x_j}\right)\right) \quad (8)$$

where $p\left(k \mid \boldsymbol{x_i}\right)$ is the posterior probability of that sample $\boldsymbol{x_i}$ is generated by the *k*-th component in GMM. And the same as the equation (3), if $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ are in the same class, $w_{ij} = 1/N_+$, otherwise, $w_{ij} = 1/N_-$. Thus, a local metric is learned by minimizing equation (9) with the PSD constrain of $\boldsymbol{M_k}$.

$$F_k\left(\boldsymbol{M_k}\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}^k f_{\boldsymbol{M_k}}\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right) \quad (9)$$

The same as the equation (3), $f_{\boldsymbol{M_k}}\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right)$ represents the log logistic loss of a sample pair, $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$. A set of local metrics, $\boldsymbol{M_k}(k = 1, 2, ..., K)$, are obtained by minimizing the equation (9) with the APG approach [28]. Then local and global metric matrixes are combined as described in the equation (6) when evaluating the similarity between testing samples. Specially, we call this approach as Local metric learning for MLAPG (L-MLAPG).

## 4. Experiments

We conduct experiments on three challenging datasets (VIPeR [1], PRID450S [10] and QUML GRID [11]) to evaluate the proposed approaches from both metric learning and feature representation perspectives. Different kinds of global metric learning methods and feature descriptors are used to demonstrate the effectiveness of the proposed framework of IGLML. Additionally, experiments of the L-MLAPG method also show more potentials.

### 4.1. Datasets and experiment protocols

The Cumulative Matching Characteristic (CMC) curve and the Rank-1 accuracy [12] are most widely used evaluation criterions of re-id. The CMC curve reflects the probabilities of finding the exact matched results in the first *r* ranks. When *r*=1, it corresponds to Rank-1 accuracy, which represents the percentage of right matched query samples in the testing set.

We validate the proposed approaches on VIPeR [1], PRID450S [10] and QMUL GRID [11] datasets. These datasets contain images of people in non-overlapping camera views. In each time of experiments, we randomly divide the dataset into two parts, half of which are used for training and the other half are used for testing. The experiment is repeated for 10 times, and the average performance is obtained.

Generally, images from one view consist of the gallery set, and ones from the other view consist of the probe set. The task of re-id is to find matched results from the gallery set

Table 1: Comparison of the proposed IGLML and global metric learning approaches with LOMO features [2] on different datasets.

| Methods | **VIPeR** [1] | | | **PRID450S** [10] | | | **QMUL GRID** [11] | | |
|---|---|---|---|---|---|---|---|---|---|
| | Rank-1 | Rank-10 | Rank-20 | Rank-1 | Rank-10 | Rank-20 | Rank-1 | Rank-10 | Rank-20 |
| IGLML-XQDA | **41.99** | **82.50** | **92.25** | **60.62** | **89.82** | **94.62** | **18.80** | **44.08** | **55.52** |
| XQDA [2] | 40.00 | 80.51 | 91.08 | 59.60 | 89.60 | 93.91 | 18.32 | **44.08** | 55.44 |
| IGLML-MLAPG | **42.47** | **83.45** | **93.29** | **59.73** | **90.44** | **95.56** | **18.08** | **43.44** | **55.92** |
| MLAPG [6] | 40.73 | 82.34 | 92.37 | 58.76 | 90.31 | 95.33 | 17.68 | 43.28 | 55.28 |

for the given query images from the probe set. We utilized the single-shot setting in our experiments.

## 4.2. Evaluations of metric learning approaches

The proposed IGLML can be used to improve the performances of many existing metric learning approaches. In the experiments, different global metric learning methods such as XQDA [2] and MLAPG [6] are integrated into the proposed framework to validate the wide effectiveness of IGLML with different global metric learning methods. LOMO [2] feature is used in the experiments.

### 4.2.1 Experiments on VIPeR

VIPeR [1] is one of the most widely used dataset in re-id. It contains 632 pairs of images from two cameras. The unified size of these images are scaled to $128\times48$ pixels. We randomly select 316 pairs for training and the other 316 pairs for testing. The experiment is repeated for 10 times to get an average performance.

Table 1 shows the accuracy of Rank-1, 10, 20 of the proposed approaches comparing to the corresponding global metric learning methods on VIPeR. The Rank-1 accuracy of IGLML based on XQDA is 41.99%, which raises the performance by 1.99% comparing to XQDA [2]. And the Rank-1 accuracy of IGLML based on MLAPG is 42.47%, which is 1.74% higher than that of MLAPG [6].

The parameters of the proposed IGLML method are selected by empirical evaluation. The experimental results show that our work achieves better performance than global metric learning approaches in various degrees when the numbers of GMM components ($K$) and PCA components ($D$) are in the scopes of **[2,10]** and **[10,50]** or larger scopes, respectively. For the results of VIPeR in Table 1, the number of GMM components, $K$ =4, the reduced feature dimension after PCA, $D$ =27, and the weight of global metric matrix in the equation (6), $\alpha_0 = 2$ .

### 4.2.2 Experiments on PRID450S

PRID450S [10] contains 450 pairs of images in two camera views. The size of each image is not unified. In order to extract LOMO features [2], we resize the images into $128\times64$ pixels. And we use the code provided by [2] to compute LOMO features which are 26,960 dimensions. Then XQDA, MLAPG and IGLML based on them are realized with the extracted LOMO features.

The middle columns of Table 1 show the accuracy of Rank-1, 10, 20 of the related approaches. The Rank-1 accuracy is improved by 1.02% comparing to XQDA which arrives at 60.62% and 0.97% comparing to MLAPG.

Since the number of samples in PRID450S is less than VIPeR dataset, the parameters of GMM components might be lower. For the results of PRID450S shown in Table 1, the parameters are set as following: $K$ =3, $D$ =30, $\alpha_0 = 2$ .

### 4.2.3 Experiments on QMUL GRID

The QMUL underGround Re-identification (GRID) dataset [11] contains 250 image pairs of persons captured form cameras in an underground station. Additionally, there are 775 images of persons who are not in the 250 pairs. We randomly select 125 pairs of images for training. The other 125 pairs and the 775 non-labeled images construct the testing set. The experiment is repeated for 10 times to obtain average performance.

Since the number of training samples is very limited in the GRID dataset. When we use GMM clustering and the partitioning strategy described in Section 3.2, there will be only a few samples in each local subset. However, IGLML needs both enough GMM components to learn different local characteristics adequately and enough training samples in each subset to avoid over-fitting for localized learning. Thus, features' dimensions after PCA might be lower and each element value in the covariance matrixes of GMM components are enlarged 10 times to obtain smoother posterior probabilities considering of the matrix inversion operator. Then if a sample's posterior probability of the $k$-th component is larger than 0.01, the sample will be partitioned into the local subsets corresponding to the $k$-th component. These skills make lager overlapping between subsets to ensure both enough GMM components and training samples in each subset.

The columns on the right of Table 1 show Rank-1, 10, 20 accuracy of IGLML comparing to the corresponding global learning methods. We replicated the experiments by the source codes provided by [2] and [6]. The proposed approach is realized with the same experiment setting with XQDA and MLAPG in our experiments. As Table 1 shows, IGLML approaches improve the performance a little.

Parameters for the results of IGLML based on XQDA are set as $K$=4, $D$=16, $\alpha_0 = 2$. And IGLML based on MLAPG shares the parameters as $K$=5, $D$=12, $\alpha_0 = 2$.

#### 4.2.4 Analysis of results and parameters

The experimental results show that the proposed IGLML approach is especially suitable for datasets with complex variations of backgrounds or clothes. IGLML performs better than the global metric learning approaches on the challenging VIPeR and PRID450S datasets. For GRID dataset, the proposed methods improve the performance a little, which is not as obviously as the other two datasets. The main reason is that comparing to VIPeR and PRID450S, person snapshot images in GRID share similar background environments in an underground station. In IGLML, GMM is used to divide the training set before localized learning. It assigns images with similar backgrounds or clothes to the same local set. However, it's difficult to partition images in GRID with all similar appearances. From the experimental results, we conclude that the proposed IGLML is better at processing more challenging re-id tasks with complicated variations of backgrounds or clothes in a camera view.

The parameters of the dimensions after PCA reduction ($D$), component number of GMM ($K$), and the weight of global metric ($\alpha_0$) are selected by cross-validation. We analyze the influence of varying $K$ in section 3.3. And experimental results show that $D$ has less influence on the accuracy than $K$. IGLML performs well when $D$ varying in the range of **[20, 50]** or a larger range on VIPeR and PRID450S datasets.

### 4.3. Evaluations of local metric learning

We also analyze the performance of purely local metric learning, which means $\alpha_0 = 0$ in the equation (6). We conduct experiments of L-MLAPG and local metric learning based on XQDA and MLAPG appoaches on VIPeR dataset with LOMO feature [2]. Local metrics are learned with the same parameters, where $D$=41, $K$=5.

Table 2 shows the accuracy of the related methods. It can be seen that L-MLAPG performs better than other local metric learning methods. IGLML doesn't perform as well as global methods in [2] and [6] when $\alpha_0 = 0$. Local metrics in IGLML are learned on only a part of the training set. The generalization ability of the models is not as strong as that learned from the entire training set. However, L-MLAPG learns local metrics on all of the training samples by adapting the weights of loss functions. It achieves the Rank-1 accuracy of 41.39%, which is higher than that of MLAPG [6] and other local metric learning approaches. If introducing global metric into the integrated metric by adjusting $\alpha_0$ in L-MLAPG, it will perform much better.

Table 2: Comparison of purely local metric learning of IGLML and L-MLAPG on VIPeR dataset [1] with LOMO features [2].

| Methods | Rank-1 | Rank-10 | Rank-20 |
|---|---|---|---|
| L-MLAPG | **41.39** | **83.13** | **93.39** |
| IGLML-MLAPG( $\alpha_0 = 0$ ) | 40.16 | 81.84 | 91.65 |
| IGLML-XQDA( $\alpha_0 = 0$ ) | 37.78 | 79.30 | 90.13 |
| MLAPG [6] | 40.73 | 82.34 | 92.37 |
| XQDA [2] | 40.00 | 80.51 | 91.08 |

Table 3: The performance of IGLML based on XQDA with different features on VIPeR dataset [1].

| Methods | $r$=1 | $r$ =10 | $r$ =20 |
|---|---|---|---|
| LOMO[2]+IGLML | **41.99** | **82.50** | **92.25** |
| LOMO[2]+XQDA[2] | 40.00 | 80.51 | 91.08 |
| FFN(original)[4]+ IGLML | 31.58 | 71.80 | 83.99 |
| FFN(original) [4]+XQDA [2] | 28.86 | 68.13 | 81.14 |
| FFN(normalized) [4]+ IGLML | **32.59** | **73.86** | **86.49** |
| FFN(normalized) [4]+XQDA [2] | 30.13 | 72.75 | 85.73 |
| GOG(original) [3]+ IGLML | 42.15 | 83.67 | 91.90 |
| GOG(original) [3]+XQDA[2] | 38.77 | 81.30 | 91.36 |
| GOG(normalized) [3]+ IGLML | **43.89** | **85.16** | **93.64** |
| GOG(normalized) [3]+XQDA[2] | 42.53 | 84.40 | 92.97 |
| WHOS(original)[34]+IGLML | 40.19 | 82.06 | 90.89 |
| WHOS(original)[34]+XQDA[2] | 33.39 | 74.62 | 85.82 |
| WHOS(normalized)[34]+IGLML | **42.91** | **84.65** | **93.07** |
| WHOS(normalized)[34]+XQDA[2] | 41.61 | 82.72 | 92.82 |

### 4.4. Evaluations of feature representations

We also conduct experiments with different features to verify the generalized effectiveness of IGLML. LOMO [2], GOG [3], an enhanced deep feature extracted by FFN [4] and WHOS [34] features are used in the experiments.

#### 4.4.1 Feature representations

LOMO [2] feature is robust against illumination and viewpoints variations. Multi-scale Retinex transformation [13] is used to preprocess the images in order to overcome the color distortion caused by illumination. Then it extracts features by a set of sliding windows on the images. Color and texture histograms are computed in each sliding window. Then it computes the maximal value of each bin in the histograms among sub-windows at the same horizontal location to handle the change of viewpoints.

GOG [3] is a regional descriptor based on a hierarchical distribution of pixel features. Location, gradient and color features are extracted inside a local patch. And a Gaussian distribution of pixel features represents the appearance of a local patch. Then the characteristics of the patches in a lager region are described by another Gaussian distribution. We use GOG descriptor in RGB space in the experiments.

Table 4: Comparison of state-of-the-art results on VIPeR [1].

| Method | | *r*=1 | *r* =10 | *r* =20 |
|---|---|---|---|---|
| Proposed | IGLML-MLAPG | **42.47** | 83.45 | 93.29 |
| | L-MLAPG | 41.39 | 83.13 | **93.39** |
| | IGLML-XQDA | 41.99 | 82.50 | 92.25 |
| State-of-the-art | MLAPG[6] | 40.73 | 82.34 | 92.37 |
| | XQDA[2] | 40.00 | 80.51 | 91.08 |
| | KEPLER [30] | 42.41 | 82.37 | 90.70 |
| | RMLLC [31] | 31.27 | 75.31 | 86.71 |
| | SCNCD [33] | 37.80 | 81.20 | 90.40 |
| | KCCA [26] | 37.00 | **85.00** | 93.00 |
| | LFDA [32] | 24.18 | 67.12 | 78.96 |
| | KISSME [5] | 19.60 | 62.20 | 77.00 |

Table 5: Comparison of state-of-the-art results on PRID450S [10].

| Method | | *r*=1 | *r* =10 | *r* =20 |
|---|---|---|---|---|
| Proposed | IGLML-MLAPG | 59.73 | 90.44 | 95.56 |
| | L-MLAPG | 59.56 | **90.62** | **95.96** |
| | IGLML-XQDA | **60.62** | 89.82 | 94.62 |
| State-of-the-art | MLAPG[6] | 58.76 | 90.31 | 95.33 |
| | XQDA[2] | 59.60 | 89.60 | 93.91 |
| | Mirror KMFA [8] | 55.42 | 87.82 | 93.87 |
| | LFDA [32] | 36.18 | 72.40 | 82.67 |
| | KISSME [5] | 36.31 | 75.42 | 83.69 |

Enhanced deep feature extracted by FFN [4] combines hand-crafted and CNN features [29] effectively. It jointly maps ELF16 [8] and CNN features to a unitary space. The parameters of the CNN net are influenced by hand-crafted features by back propagation.

WHOS [34] splits the image into several overlapping horizontal stripes. Weighted histograms in HSV, RGB and Lab color spaces are extracted. The weight of each pixel in the histogram bins is computed by a non-isotropic Gaussian kernel centered in the image to decrease the influence of background information. In addition, texture descriptors such as LBP and HOG are also extracted. All of the color and texture histograms are concatenated to obtain a more discriminative feature representation.

#### 4.4.2 Experiments with different features

LOMO features are already normalized, which can be used directly, but GOG, FNN and WHOS features are not normalized. Actually, for high-dimensional features, normalization is important to improve the performance [23]. We do experiments on both original and normalized feature to observe the improvements of IGLML. We used $L_2$ norm as described in [3] to normalize the features.

Table 3 shows the performance of IGLML based on XQDA comparing to XQDA with different feature representations. For GOG, FFN and WHOS, the results of both original non-normalized and normalized features are shown. It can be seen that IGLML can improve the performance of global metric learning no matter what features are used. And the proposed approach can improve the performance more greatly with non-normalized features. IGLML performs better than global learning methods using both features with background influence decrease and features containing background information. It means the proposed approaches don't depend on background.

### 4.5. Comparison with state-of-the-art results

We compare the results of the proposed approaches with state-of-the-art methods. The proposed models are based on LOMO feature. Table 4 and 5 summarize the performance of some recently proposed models on VIPeR and PRID450S datasets respectively. It shows that the proposed L-MLAPG and IGLML approaches with LOMO features perform better than other listed methods generally. IGLML based on MLAPG owns the highest Rank-1 accuracy of 42.47% on VIPeR among the methods in Table 4. IGLML based on XQDA achieves Rank-1 accuracy of 60.62% on PRID450S, which is the highest in Table 5. And L-MLAPG performs better on other ranks.

## 5. Conclusion

We propose a flexible framework of Integrated Global-Local Metric Learning (IGLML). The proposed framework can generally improve the performance of global metric learning approaches regardless of using what feature representation or metric learning methods. And it's especially suitable for handling heterogeneous datasets with complex variations of backgrounds or clothes in a camera view. In this paper, we integrate some simple but effective global metric learning methods by softly dividing the training set with GMM and a proposed strategy of softly partitioning overlapping subsets. Specially, we also propose a more effective local learning approach for MLAPG (L-MLAPG) by modifying the weights of sample pairs' loss with posterior probabilities of GMM components. Actually, except for XQDA and MLAPG, other global metric learning methods can also be integrated into the proposed framework and achieve better performance. The proposed approaches avoid complicated procedure of solving optimization problems in other local metric learning methods and effectively improve the performance on person re-identification issue. The experiments on the three challenging datasets verify the generalized effectiveness of the proposed approaches.

## Acknowledgements

This work is supported by National Natural Science Foundation of China (NSFC, No. 61273285, No. 61375019, No. 61673269)

## References


[1] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *IEEE PETS Workshop*, 2007.

[2] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person re-identification by local maximal occurrence representation and metric learning. In *IEEE CVPR,* 2015.

[3] T. Matsukawa, T. Okabe, E. Suzuki and Y. Sato. Hierarchical Gaussian descriptor for person re-identification. In *IEEE CVPR*, 2016.

[4] S. Wu, Y.-C. Chen, X. Li and A.-C. Wu. An enhanced deep feature representation for person re-identification. In *IEEE WACV*, 2016.

[5] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *IEEE CVPR*, 2012.

[6] S. Liao and S. Z. Li. Efficient psd constrained asymmetric metric learning for person re-identification. In *IEEE ICCV*, 2015

[7] J. Bohné, Y. Ying, S. Gentric, and M. Pontil. Large margin local metric learning. In *ECCV*, 2014

[8] Y.-C. Chen, W.-S. Zheng, and J. Lai. Mirror representation for modeling view-specific transform in person re-identification. In *IJCAI*, 2015.

[9] B. Moghaddam, T. Jebara, and A. Pentland. Bayesian face recognition. Pattern Recognition, 33(11):1771–1782, 2000.

[10] P. M. Roth, M. Hirzer, M. Kostinger, C. Beleznai, and H. Bischof. Mahalanobis distance learning for person reidentification. In Person Re-Identification, pages 247–267. 2014.

[11] C. C. Loy, T. Xiang, and S. Gong. Multi-camera activity correlation analysis. In *IEEE CVPR,* 2009.

[12] F. Porikli and A. Divakaran. Multi-camera calibration, object tracking and query generation. *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference on. IEEE, 2003*, 1: I-653-6 vol. 1.

[13] D. J. Jobson, G. Rahman and G. A. Woodell. A mul-tiscale retinex for bridging the gap between color images and the human observation of scenes[J]. IEEE Transactions on Image processing, 1997, 6(7): 965-976.

[14] D. Gray, H. Tao. Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features. In *ECCV*, 2008.

[15] R. Zhao, W. Ouyang, and X. Wang. Person re-identification by salience matching. In *ICCV*, 2013.

[16] W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In *IEEE CVPR,* 2011.

[17] W. Li, X. Wang. Locally Aligned Feature Transforms across Views. In *IEEE CVPR*, 2013.

[18] S. Saxena and J. Verbeek. Coordinated Local Metric Learning. In *IEEE International Conference on Computer Vision Workshop (ICCVW).* 2015.

[19] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *IEEE CVPR*. 2015.

[20] I. Giotis and N. Petkov. Cluster-based adaptive metric classification[J]. Neurocomputing, 2012, 81: 33-40.

[21] S. Wold, K. Esbensen, and P. Geladi. Principal component analysis[J]. Chemometrics and intelligent laboratory systems, 1987, 2(1-3): 37-52.

[22] S. Huang, J. Lu, J. Zhou and A K. Jain. Nonlinear Local Metric Learning for Person Re-identification[J]. arXiv:1511.05169, 2015.

[23] J. Sanchez, F. Perronnin, T. Mensink, and J. J. Verbeek. Image classification with the Fisher vector: Theory and practice[J]. International Journal of Computer Vision, 105(3): 222-245, 2013.

[24] D.-C. Zhan, M. Li, Y.-F. Li, and Z.-H. Zhou. Learning instance specific distances using metric propagation. In *ICML*, 2009.

[25] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. Proceedings of the 18th ACM international conference on Multimedia. ACM, 2010: 1469-1472.

[26] G. Lisanti, I. Masi and A. Del Bimbo. Matching people across camera views using kernel canonical correlation analysis. In Proceedings of the International Conference on Distributed Smart Cameras. ACM, 2014: 10.

[27] S. Calinon, F. Guenter and A. Billard. On learning, representing, and generalizing a task in a humanoid robot[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2007, 37(2): 286-298.

[28] P. Tseng. On accelerated proximal gradient methods for convex-concave optimization [J]. submitted to SIAM Journal on Optimization, 2008.

[29] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev and J. Long. Caffe: Convolutional architecture for fast feature embedding. In Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014: 675-678.

[30] N. Martinel, C. Micheloni, G L. Foresti. Kernelized saliency-based person re-identification through multiple metric learning[J]. IEEE Transactions on Image Processing, 2015, 24(12): 5645-5658.

[31] J. Chen, Z. Zhang and Y. Wang. Relevance metric learning for person re-identification by exploiting listwise similarities[J]. IEEE Transactions on Image Processing, 2015, 24(12): 4741-4755.

[32] S. Pedagadi, J. Orwell, S. Velastinand B. Boghossian. Local fisher discriminant analysis for pedestrian re-identification In *IEEE CVPR*, 2013.

[33] Y. Yang, J. Yang, J. Yan S. Liao and S.-Z. Li. Salient color names for person re-identification. In *ECCV*, 2014.

[34] G. Lisanti, I. Masi, A. D. Bagdanov and A. Del Bimbo. Person re-Identification by iterative re-weighted sparse ranking [J]. IEEE Transactions on PAMI, 2015, 37(8):1629-1642.